

Application of Advanced Statistical Techniques for Horticultural Crops Research: Present Status and Future Prospects

R Venugopalan

ICAR-Indian Institute of Horticultural Research, Bengaluru

Received 09 February 2026; Revised 13 April 2026; Accepted 14 April 2026

SUMMARY

Statistical methods play a vital role in different areas of horticultural crops research right from planning of any research experiment till selection of appropriate tool for data analysis to unravel many of the hidden results. They are central to horticultural research, enabling robust analysis of complex research issues entailing multi-disciplinary fields such as crop improvement, production, protection, post-harvest management, climate resilience, and socio-economic studies. The present communication, while delineating several of such multidisciplinary research outcomes, it also provides a road map for its potential application in emerging areas of research. Further, detailed descriptions of the statistical methodologies employed are referenced to the original sources for readers seeking methodological specifics.

Keywords: Carbon Sequestration; Crop Modelling; Crop-varietal release; Horticulture; Machine learning models; Mass transfer kinetics.

1. INTRODUCTION

Horticulture occupies a distinctive position within agricultural research due to its extensive reliance on advanced statistical methodologies. This reliance arises from the need to evaluate treatment effects in field experiments involving horticultural crops with statistical rigor and scientific validity. Given the biological variability inherent in horticultural systems, appropriate experimental design and robust statistical analysis are essential for drawing reliable conclusions.

For more than six decades, statistical experts at the ICAR–Indian Institute of Horticultural Research (ICAR-IIHR), Bengaluru, have provided critical research support to scientists and students. This multi-disciplinary research approach has encompassed the formulation of suitable experimental designs, application of appropriate statistical techniques, and interpretation of results across a wide range of horticultural research domains. These domains include fruits, vegetables, ornamental crops, and medicinal and

aromatic plants, with applications in crop improvement, production, pathology, and entomology.

The present communication elucidates outcomes derived from comprehensive statistical guidance provided to multi-disciplinary research studies in crop improvement, crop production, crop protection, natural resource management, and post-harvest technology, with an emphasis on practical applicability. Detailed descriptions of the statistical methodologies employed are referenced to the original sources for readers seeking methodological specifics. Future direction in all these aspects are also delineated individually by targeting the objectives of the emerging research problems.

2. CROP IMPROVEMENT

2.1 Parametric vs. Non-parametric stability analysis

Plant breeding is a central component of horticultural research, contributing to the development of new varieties with improved yield, resistance to biotic stresses, enhanced quality attributes, and adaptability

to diverse environments. In horticultural crops, both conventional and modern breeding approaches are employed, while Plant Variety Protection (PVP) frameworks safeguard the rights of breeders and farmers and promote the sustainable use of genetic resources. As part of varietal development, stability models are commonly used to identify promising lines for release at the conclusion of multi-year/location/season evaluation trials. However, conventional approaches often identify different stable lines for individual traits, complicating holistic varietal recommendation. To overcome this limitation, nonparametric stability models based on a trait pyramiding concept were developed to identify lines stable across all evaluated traits.

Traditional parametric methods, particularly analysis of variance, are widely employed in breeding programs to assess genotype stability. However, in practice, the relative ranking of genotypes across environments is often of greater importance than the absolute magnitude of variation. Consequently, nonparametric, rank-based approaches provide a robust and effective alternative for evaluating the consistency and stability of crop variety performance across diverse environments.

2.2 Nonparametric approach of stability analysis

Several nonparametric indices have been developed to evaluate yield stability (Thennarasu, 1995; Nassar and Huhn, 1987). These methods rely on the relative ranking of genotypes across the tested environments. The ranking is based on values of Y_{ij} with lowest Y_{ij} value receiving the rank 1, the next higher value 2 and so on. The non-parametric stability statistics derived from these genotype yield ranks across environments are described here.

$$NP_i^{(2)} = \frac{1}{n} \left[\sum_{j=1}^n |r_{ij}^* - Md_i^*| / Md_i \right]$$

$$NP_i^{(1)} = \frac{1}{n} \sum_{j=1}^n |r_{ij}^* - Md_i^*|$$

$$NP_i^{(3)} = \frac{\sqrt{\sum (r_{ij}^* - \bar{r}_{i.})^2 / n}}{\bar{r}_{i.}}$$

$$S_i^{(1)} = 2 \sum_{j=1}^{n-1} \sum_{j'=j+1}^n |r_{ij} - r_{ij'}| / [n(n-1)]$$

$$S_i^{(2)} = \sum_{j=1}^n (r_{ij} - \bar{r}_{i.})^2 / (n-1)$$

$$Si(4) = \sqrt{\sum_{j=1}^n \frac{(r_{ij} - \bar{r}_{i.})^2}{n}}$$

The rank r_{ij} is determined based on the rank of i^{th} genotype in j^{th} environment (Y_{ij}). The uncorrected Y_{ij} has the drawback that they may show significance even when there is no GE interaction. Hence, rank r_{ij}^* is determined based on corrected phenotypic values $Y^*_{ij} = (Y_{ij} - \bar{Y}_{i.}) / (\bar{Y}_{i.})$, $(\bar{Y}_{i.})$ being the mean performance of i^{th} genotype. The corrected values depend only on the GE interaction and error components. Mdi^* is the median ranks for adjusted values. These measures are widely used to assess the stability for different characters **individually** in any crop improvement research.

2.3 Non-parametric method for crop variety release developed at ICAR-IIHR

Existing nonparametric crop stability methods derive statistical indices based on genotype rankings, typically using mean or median ranks, with performance evaluated on a trait-by-trait basis. However, plant breeders often prioritize traits that are absent or inadequately expressed in existing varieties, and the arbitrary weighting of such traits can introduce bias into selection decisions. In practice, breeders seek genotypes that exhibit consistent performance across multiple traits, environments, and seasons, rather than superiority in only a limited number of traits.

To address these limitations, a stability-based weighting approach was developed that accommodates both favorable and unfavorable traits within a unified framework. This approach led to the formulation of a nonparametric stability index, termed the *Venugopalan Index* (Venugopalan *et al.*, 2020), which quantifies genotype contributions to genotype $\times$ environment (GE) interactions based on relative performance and multi-trait stability, with demonstrated application in okra (*Abelmoschus esculentus* L.) breeding. A step-by-step process is depicted in a flowchart.



Fig. 1. Schematic representation of the non-parametric approach for crop variety release developed at ICAR-IIHR

The Venugopalan nonparametric index was further standardized to identify stable lines for crop variety release by simultaneously considering stability and combined performance across multiple traits. The indices were evaluated and validated using experimental data from okra, dolichos, cowpea, French bean, China aster, chrysanthemum, banana, and guava.

Okra: A nonparametric stability index was developed to support crop varietal release by integrating multi-trait performance with stability across years. Eight okra hybrids were evaluated over three consecutive years for yield, flowering, plant height, and disease incidence. Rank-based nonparametric measures were applied to assess genotype stability, thereby minimizing sensitivity to outliers and accommodating nonlinear genotype $\times$ environment interactions. A composite index weighted according to trait importance was used to identify genotypes combining superior performance with stability. Based on the composite index, the hybrids OKMSH-2, OKMSH-3, and OKMSH-7 were identified as superior, in that order. OKMSH-2 recorded the highest nonparametric value (1223.67) across all evaluated traits. Its superior ranking was largely attributed to the higher weight assigned to yellow vein mosaic virus incidence, for which it exhibited the lowest disease severity, along with high

yield performance. The proposed approach enables breeders to select genotypes with consistent multi-trait performance and is adaptable to multi-location trials, providing a practical tool for varietal recommendation.

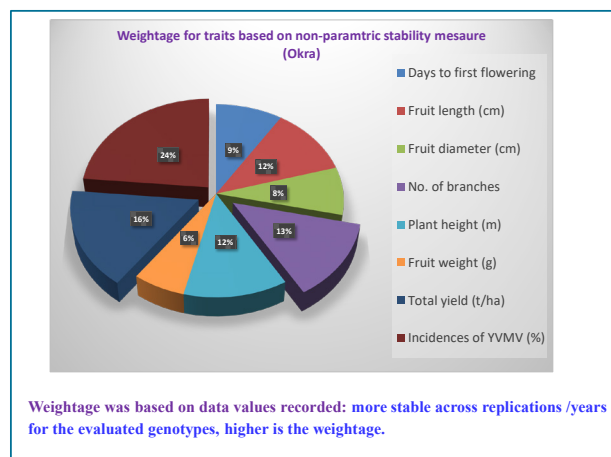


Fig. 2. Weightage based on data values recorded in Okra

Details can be perused at (Venugopalan *et al.* 2020)

China aster: A nonparametric ranking approach was developed for China aster to identify stable and superior lines based on overall multi-trait performance across growing seasons. Trait importance was determined using stability-based weights derived from consistency across years and within replications, avoiding arbitrary weighting and accommodating traits where either increase or decrease is desirable. The approach also accounts for traits where either increase or decrease is desirable relative to existing cultivars. Seven stability parameters (Si_1 , Si_2 , Si_4 , NPi_1 , NPi_3 , and the IIHR method) were applied to evaluate seven traits: plant height, plant spread, number of branches per plant, flower diameter, number of flowers per plant, stalk length, and vase life over three years (2021–2023). Cluster analysis grouped the genotypes into three distinct clusters. Based on the combined nonparametric stability index computed for 36 lines, genotypes 15-41-3, followed by 15-41-1 and 15-41-16, were identified as the most stable across all traits. These lines are recommended for use in hybridization and varietal improvement programs, particularly under multi-location trial conditions. The system offers a more holistic approach to variety assessment and selection.

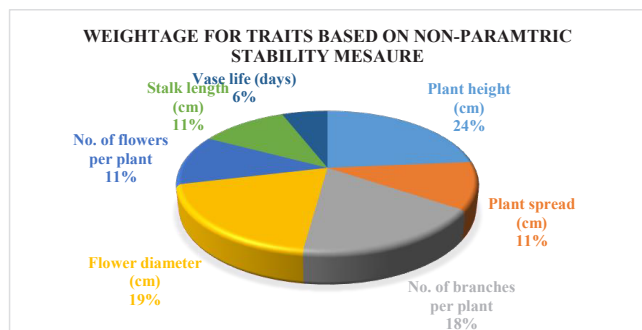


Fig. 3. The weightage of various traits computed (for combined non parametric index) China aster

Details can be perused at (Venugopalan *et al.* 2025)

Chrysanthemum: Through trait pyramiding concept, non-parametric stability models are developed to identify stable line(s) across yield and related traits. This concept is demonstrated using various yield attributing traits of 15 chrysanthemum lines evaluated under AICRP on Floriculture at ICAR-Indian Institute of Horticultural Research, Bengaluru during 2023 to 2025. R-codes were developed for the analysis. Based on the Non-parametric stability models of venugopalan indices, it was found that IIHR 4-8 and IIHR 2-16 are stable across all traits evaluated (Flower diameter (cm), number of flowers/plant, flower stalk length (cm), plant height (cm), plant spread (cm). These stable lines will be utilized effectively to release a variety holistically combining all desirable traits in crop hybridization trials, emphasizes the need to construct a non-parametric index for assessing the stability of a set of lines collectively based on various traits evaluated over multiple seasons or years in an experimental setup.

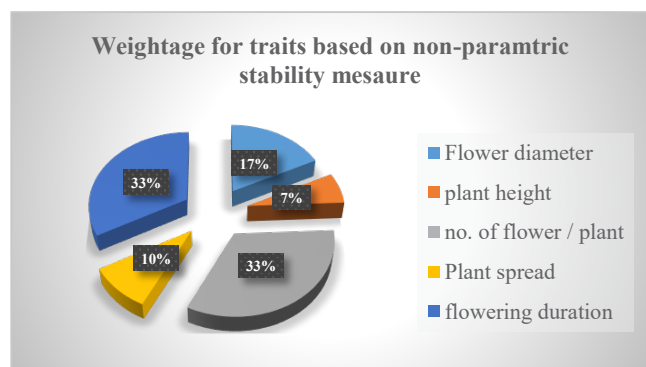


Fig. 4. The weightage of various traits computed (for combined nonparametric index) Chrysanthemum

Details can be perused at (IIHR Annual Report, 2025)

2.4 Future directions in crop stability analysis research

Future research both in parametric and nonparametric stability analysis is expected to focus on integrating rank-based methods with advanced statistical and computational approaches to better capture complex genotype $\times$ environment ($G \times E$) interactions. Hybrid frameworks combining nonparametric indices with mixed models and Bayesian approaches can enhance inference by incorporating random effects, prior information, and uncertainty estimation while retaining robustness to non-normality and outliers.

The extension of nonparametric stability measures to high-dimensional, multi-trait datasets using machine learning techniques such as random forests, gradient boosting, and clustering algorithms represents another promising direction. These methods can identify nonlinear relationships and trait interactions that traditional models may overlook. Additionally, multi-environment and multi-season stability assessment can be strengthened by functional data analysis and resampling-based techniques, enabling dynamic stability evaluation over time. Future work should also emphasize the integration of nonparametric stability indices with genomic selection and phenomics data, allowing early-stage prediction of stable genotypes for effective use in crop hybridization trials, thereby hastening the breeding cycle simultaneously by fixing multiple traits. Selection based on a multidimensional scaling approach, as adopted in speed breeding programmes, can further strengthen breeding strategies by reducing the breeding cycle and enabling the simultaneous fixation of multiple desirable traits through a trait-pyramiding concept.

2.5 Incomplete block designs in germplasm evaluation trials

Randomized Block Design (RBD) is widely used for evaluating germplasm accessions along with local checks; however, when the number of test treatments is large and seed availability is limited, alternative designs are required. Balanced Incomplete Block Design (BIBD), where all treatments do not appear in any block, offers greater precision while conserving resources. This approach reduces seed requirement and input costs such as land, labour, water, and agrochemicals.

Balanced incomplete block design structure was developed for evaluating breeding germplasm lines as a measure of improving input use efficiency in crop research resulted in saving 60.2 % of the land area required for conducting germplasm evaluation in Okra, calls for application in other crops wherever input use efficiency is required to be enhanced.

Details can be perused at (Venugopalan *et al.* 2008.)

2.6 Future prospects in use of Incomplete Block Designs for Germplasm Evaluation

The incorporation of genomic information into IBD analysis through genomic best linear unbiased prediction (GBLUP) and Bayesian genomic models represents a major future direction, allowing early and more accurate prediction of germplasm performance. Machine learning approaches, including ensemble learning and Gaussian process models, can complement traditional analyses by capturing nonlinear genotype $\times$ environment interactions, focussing on adaptive and optimal design strategies.

3. CONSTRUCTION OF SELECTION INDICES

3.1 Selection index for identifying factors attributing to aril browning in pomegranate progenies

Genotypic variation in aril browning and its association with key fruit quality traits in 158 progenies derived from complex pomegranate hybrids using appropriate statistical modeling approaches were carried out. Significant variability was observed for aril browning severity, skin colour, aril colour, total soluble solids (TSS), and seed mellowness. Spearman's rank correlation analysis indicated a negative association between aril browning and aril colour, while TSS showed a positive association with browning severity. An optimized regression model demonstrated that aril colour and TSS together accounted for 73.5% of the total variation in aril browning. Darker aril colour was associated with reduced browning, whereas higher TSS increased susceptibility to browning. These findings suggest that effective breeding strategies aimed at reducing aril browning should balance aril colour intensity and sweetness, enabling early and reliable selection of superior progenies and the constructed selection index could be used as a criterion for early selection of future progenies.

Details can be perused at (Jalilop *et al.* 2010)

3.2 Marigold

A Smith–Hazel selection index was constructed to identify superior marigold sterile and fertile lines based on total number of flowers per plant, flower yield per plant, carotenoid content, and number of seeds per capitulum. Among the sterile lines, three superior genotypes were identified for each trait. For total number of flowers per plant and flower yield per plant, the superior sterile lines were IIHRMO 2329, IIHRMO 2340, and IIHRMO 2347. For carotenoid content, the selected sterile lines were IIHRMO 2352, IIHRMO 2340, and IIHRMO 2336, while IIHRMO 2339, IIHRMO 2340, and IIHRMO 2331 were superior for number of seeds per capitulum. Similarly, among the fertile lines, IIHRMO 41, IIHRMO 2329, and IIHRMO 2335 were identified as superior for total number of flowers per plant and flower yield per plant. For carotenoid content, the superior fertile lines were IIHRMO 237, IIHRMO 234, and IIHRMO 2335, whereas IIHRMO 2310, IIHRMO 2333, and IIHRMO 2335 exhibited superiority for number of seeds per capitulum.

3.3 Future prospects: Genomic selection models in breeding studies for maximizing progress from breeding

Knowledge of parental breeding values is critical for maximizing genetic progress, particularly in perennial crops. Breeding value represents the expected average performance of progeny derived from a parent across multiple crosses, enabling targeted hybridization for desired traits. In mango, early improvement relied on selection from natural out-crossed seedlings via open pollination, where only the female parent was identifiable. With controlled hand pollination, both seed and pollen parents can be tracked, allowing estimation of breeding values using best linear unbiased prediction (BLUP) approaches. In crop improvement trials, estimated breeding values derived from BLUP can guide the selection of parents to achieve specific objectives—for instance, producing a higher frequency of progeny with intermediate fruit size preferred in the market. Such targeted hybridization accelerates the breeding process and increases the efficiency of selection for quantitative traits.

Future directions must involve integrating genomic selection (GS) models such as G-BLUP

and Bayesian approaches with multi-trait and multi-environment datasets. Coupled with high-throughput phenotyping and molecular marker data, these models can predict superior progeny at early stages, accelerate breeding cycles, and optimize parent selection for targeted traits, thereby enhancing the precision and efficiency of perennial crop improvement programs. Statistical models constructed with inputs from ANN, CNN, GNN approaches would capture the nonlinearity present in the system effectively as compared to their linear counterpart models.

3.4 Genome-wide association study identifies key quantitative trait loci (QTL) for fruit morphometric traits in perennials

Future research should validate the identified GWAS loci across diverse populations and environments to confirm their stability and transferability. Fine mapping and functional characterization of candidate genes are needed to clarify their biological roles in trait expression. Integrating GWAS results with transcriptomics, metabolomics, or epigenetic data could provide deeper insights into the underlying regulatory mechanisms. Expanding analyses to include genotype $\times$ environment interactions will improve trait prediction under varying climatic conditions. Finally, incorporating validated markers into marker-assisted or genomic selection programs could accelerate crop breeding for improved yield and quality traits.

4. CROP PRODUCTION

4.1 Robust ANOVA based method for handling multiple outliers in designed horticultural crops experiments

Classical analysis of variance (ANOVA), routinely employed for treatment comparisons in perennial crop experiments, is based on stringent assumptions of normality and homogeneity of variances. In practice, these assumptions are often violated due to the presence of outliers in replicated observations. Such outliers adversely affect data normality, inflate the coefficient of variation, and may lead to misleading or non-significant treatment effects. Although elimination of outliers may improve adherence to model assumptions, it is statistically undesirable as it violates experimental principles and results in loss of valuable information, especially under limited replication.

Robust ANOVA methods provide an effective alternative by down-weighting the influence of outliers and yielding more reliable treatment comparisons (Paul & Bhar, 2011; Venugopalan & Manjunath, 2019). This approach is particularly relevant in perennial fruit crop trials where traits such as yield per tree are recorded over multiple years following treatment imposition. In such situations, pooled ANOVA often fails to detect treatment effects due to high inter-annual variability. Robust ANOVA facilitates more realistic inference by accommodating data irregularities and temporal variability, thereby improving the reliability of conclusions.

Mango: A statistical method based on Robust ANOVA to handle outliers induced high coefficient of variation (CV) in pooled (2011-2018) analysis of long-term Mango cv. Totapuri rootstock trial was suggested. Based on the results, it was concluded that the rootstock treatment T3: Olour (average yield over the period 2011 to 2018 as 57.21 kg/tree) as the best. Precision gained as estimated by reduction in CV (%) was in the range of 11.01 % to 78.9 %. SAS IML codes were built-in for analysis. Hence, this study calls for employing robust ANOVA approach in testing the significance of evaluated treatments in a designed perennial crop experiment with high CV, alternate to Pooled ANOVA, that would have reduced the sensitivity of testing the significance of treatment differences otherwise.

Details can be perused at (Venugopalan *et al.* 2023)

Papaya: Experimental data on different yield and associated characters of Papaya (Cv Red Lady), for twelve treatments with three replications evaluated in RCBD set up in the experimental plot of ICAR-Indian Institute of Horticultural Research, Bengaluru were utilized. Results fortified the efficacy of robust methods while dealing with outliers as revealed by the three to five-fold reduction in error sum of squares coupled with acceptable probability values for all the characters. Thus this study calls for adoption of robust ANOVA approach while dealing with outliers in perennial horticulture crop experiments in future research and suggested to develop similar robust approach for carrying out pooled (over years/seasons/locations) ANOVA based real life experimental data, especially in long term perennial crop experiments, not only based on RCBD set up but also based on other design set up.

Details can be perused at (Venugopalan *et al.* 2019)

a) Integrated MANOVA statistical approach to identify best stage of harvest and post-harvest ripening of fruits treatments simultaneously based on seed yield, related traits in Pumpkin (*Cucurbita moschata*)

Pumpkin (*Cucurbita* spp., $2n = 40$) is an important cucurbit in which seed yield and quality are influenced by fruit maturity and post-harvest ripening. Seed development in cucurbits may continue after harvest due to nutrient translocation from pulp to seeds, necessitating optimization of harvest stage and ripening duration. In this study, Multivariate Analysis of Variance (MANOVA) and Principal Component Analysis (PCA) were employed (by effectively capturing the association among traits) to evaluate the combined effects of harvest stage and post-harvest ripening on seven seed yield traits in pumpkin (*Cucurbita moschata*). Fruits harvested at 45 days after pollination and ripened for 30 days emerged as the most effective treatment, demonstrating the advantage of multivariate approaches over univariate analyses which has resulted in identifying various treatments as superior for different traits. This integrated statistical approach based on linear combination of all the traits, contrary to individual trait based ANOVA study, highlights the potential to optimize agricultural practices for improved seed yield and quality in pumpkin cultivation.

Details can be perused at (R Sukanya *et al.* 2025)

C) Deep learning statistical models

Artificial Neural Network (ANN) models for identifying significant crop-logging parameters influencing crop yield across phenological growth stages in Papaya, Banana and Brinjal

Banana: Biological organisms generally exhibit nonlinear growth, unlike the linear assumptions underlying many data analysis methods. Using primary data collected from farmers' fields during 2010–12, artificial neural network (ANN) models were developed for banana (*Musa × paradisiaca* L.) and compared with multiple linear regression (MLR). ANN models more effectively captured inherent nonlinearity, achieving 8.7–11.3% higher R^2 values. Significant crop-logging traits influencing yield prediction (82.3–92.1%) included

plant girth (27 cm) and leaf breadth (35.5 cm) during the 3rd month; plant height (82 cm) and 13 leaves during the 7th month; and plant height (127 cm) and leaf breadth (54 cm) during the 9th month. Overall, ANN approaches are better suited for crop modeling due to their ability to handle nonlinear relationships among biometrical traits. These optimized traits provide valuable inputs for precision farming and sensor-based fertigation scheduling.

Details can be perused at (Venugopalan *et al.* 2010)

Brinjal: An attempt was made to assess the influence of various biometrical characters across different growth stages in Brinjal crop yield along with various yield attributing characters across four growth stages using ANN models. Results of seedling stage had indicated that plant height, girth and number of primary branches could together predict the crop yield to an extent of 83 % for training set and 62 % for validation of model accuracy. Normalized importance was 0.225, 0.142, and 0.131 respectively. In case of Vegetative stage plant height, girth and number of leaves could together predict the crop yield to an extent of 89 %, for training set and 85 % for validation of model accuracy and Normalized importance was 0.126, 0.098 and 0.133, respectively. Where as in Flowering stage both the plant spreads (east-west and north-south) could together predict the crop yield to an extent of 88 % for training set and 64 % for validation of model accuracy and Normalized Importance was 0.220 and 0.245 respectively. Finally, for fruiting stage plant height, number of primary branches and plant spread (east-west) could together predict the crop yield to an extent of 68 % for training set and 64 % model accuracy. Normalized importance was 0.229, 0.134 and 0.227 respectively. ANN captured the inherent nonlinearity among biometrical variables to predict the yield 17 % more accurately across stages. These optimized traits provide valuable inputs for precision farming and sensor-based fertigation scheduling. Also, plant yield could be maximized during early stages itself by following optimum protocols based on model predicted optimum values of crucial traits.

Crop stage-wise important biometrical factors responsible for crop yield in Brinjal

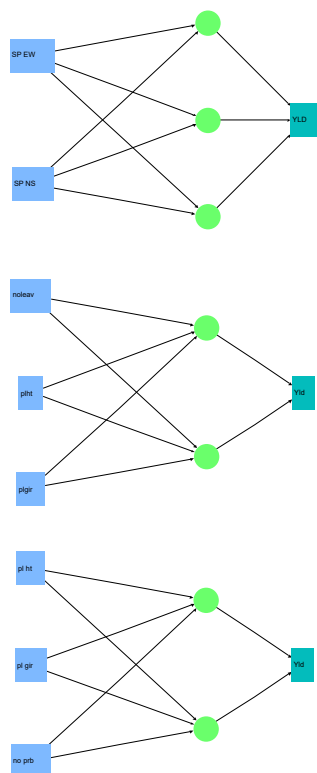


Fig. 5. Optimized ANN for a) seedling stage b) Vegetative stage and c) Flowering stage

Note: pl ht: Plant height (cm); pl gir: Plant girth (cm); no prb: No. of primary branches; noleav: No. of leaves; SP EW: Plant Spread (cm) (East-West); SP NS: Plant Spread (cm) (North- South)

Details can be perused at (Hanumanthaiah *et al.* 2018)

Banana (cv Ney Poovan): Venugopalan and Shamsundaran (2005) advocated the use of linear statistical modeling technique for evolving crop-logging models in Robusta Banana. Along the similar lines, ANN based non-linear statistical models were developed for evolving crop-logging technique with a view to identify biometrical variables which are best indicators of banana (cv Ney Poovan) yield across its various growth stages based on farmer's field data observed during 2003-05 & 2006-08. Efforts were also made to work out point and interval estimates of crop-logging variables with 95% probability interval. Thus, the stochastic models developed with identically and independently distributed (i.i.d) white noise, revealed that 48 cm of leaf length and 37 cm of plant height at 71 Days after planting (DAP), 64.5 cm of leaf length and 10 leaves at 144 DAP, 90 cm

of leaf length and 36 cm of Leaf breadth at 207 DAP, 13 leaves and 41 cm of leaf breadth at 260 DAP, 42.5 cm of leaf breadth and 14 leaves at 326 DAP, 51cm of plant girth at 380 DAP were the best indicators of crop yield. The prediction power of the crop-logging models were worked out be in the range of 93.6% - 96%. Second year results also revealed that 59 cm of leaf length & 25 cm of plant girth at 75 Days after planting (DAP); 74 cm of leaf length & 13 leaves at 150 DAP; 112 cm of leaf length, 36 cm of leaf breadth and 15 leaves at 240 DAP and 52 cm of plant girth and 18 leaves at 370 DAP were the best indicators of crop yield. Before making a final decision about the model adequacy, randomness assumption about the error term was statistically tested. These optimized traits provide valuable inputs for precision farming and sensor-based fertigation scheduling.

Crop stage-wise important biometrical factors responsible for crop yield in Banana

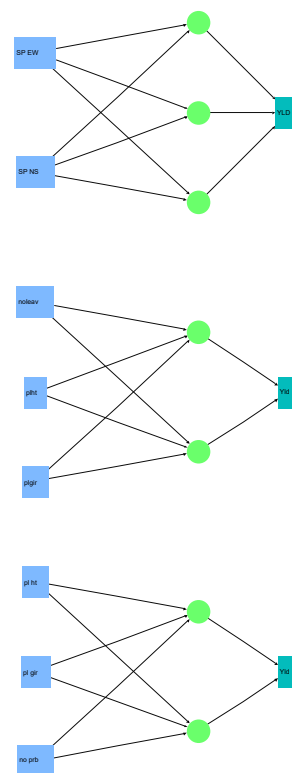


Fig. 7. Optimized ANN for a) after sucker emergence and before inflorescence stage (3rd month) b) pulp development stage (7th month) and c) starch accumulation stage (9th month)

Note: pl ht: Plant height (cm); pl gir: Plant girth (cm); no prb: No. of primary branches; noleav: No. of leaves; SP EW: Plant Spread (cm) (East-West); SP NS: Plant Spread (cm) (North- South)

ANN captured the inherent nonlinearity among biometrical variables to predict the yield having 82.3 to 92.1% power of prediction.

Details can be perused at (Venugopalan *et al.* 2010)

D) Statistical models for developing a quick method for estimation of leaf area in Fig using non-destructive sampling approach

Statistical models were developed for rapid, non-destructive estimation of leaf area in Fig. Leaf area (Y) was predicted using middle lobe length (V_2) as the explanatory variable, with the models $Y = -200.91 + 24.51V_2$ for cv. Deanna and $Y = -135.63 + 16.79V_2$ for cv. Poona, explaining 77% and 78% of the predictability, respectively. To facilitate adoption of these models in future studies, bootstrap resampling was employed to determine the optimum sample size. Resampling was performed at $n = 25-1000$, and models were refitted at each level. Based on prediction efficiency, optimal sample sizes were identified as 150 and 200. Independent validation in the subsequent year yielded R^2 values of 0.82 (Deanna) and 0.84 (Poona), confirming model robustness. Bootstrap confidence intervals were also constructed for the optimized variables. These models would serve as quick method for estimation of leaf area using non-destructive sampling there by improving the crop productivity with enhanced physiological growth.

E) Statistical models to identify the factors influencing pruning time on enhancing the yield and quality of *Jasminum sambac* flowers during off-season

The study evaluated the effect of different pruning times on inducing off-season flowering and improving yield and quality of *Jasminum sambac*. Plants were pruned at monthly intervals from August to November to assess their influence on winter flowering. Pooled results from 2005–08 showed that pruning in September significantly enhanced off-season flower yield (72.99 g/plant) and produced superior quality flower buds compared to later pruning, with November-pruned plants yielding only 25.51 g/plant of marketable buds. It was concluded using the optimized model

$Y = -61.7 + 2.635X_1 + 3.533X_9 - 1.203X_5$ ($R^2=0.81$) with maximum temperature, wind speed and stalk length as crucial factors that September pruning enhances winter flower yield and profitability due to higher seasonal prices.

Details can be perused at (Nair *et al.* 2011)

5. FUTURE PROSPECTS

5.1 ML and DL based studies

Machine Learning (ML) and Deep Learning (DL) approaches offer strong future potential for precision agriculture through accurate, early-stage yield and quality prediction. Advanced models such as ANN, Random Forest, Gradient Boosting, SVR, and LSTM can effectively capture complex nonlinear and temporal relationships among crop growth, soil, weather, and quality traits. Integration of multi-year, multi-stage crop data enables development of robust, stage-specific decision support systems. Deep learning models, particularly LSTM, are promising for real-time forecasting using sequential growth and climate data. Future studies can incorporate remote sensing, IoT, and sensor-based data to enhance model accuracy and scalability. These approaches will support site-specific management, resource optimization, and sustainable crop production.

6. CROP PROTECTION

a) Disease forecasting and Pest-population dynamics models

Development of Non-linear statistical model for disease forecasting in Grapes

Downy mildew is a major constraint to grape (*Vitis* spp.) production, causing severe yield losses under favorable weather conditions. This study evaluated the influence of climatic variables on downy mildew incidence in grape cv. Anab-e-Shahi and modelled disease progression separately for backward and fore pruning seasons. Weather-based regression models revealed that, during backward pruning, maximum temperature, evaporation, and morning relative humidity (7:30 hrs), with a one-week lag, explained 89.4% of the variation

in disease incidence. For fore pruning, minimum temperature and relative humidity at 7:30 and 13:30 hrs accounted for 88% of the variation. Disease development was favoured by temperatures of 28–32 °C (maximum) and 14–22 °C (minimum), high morning relative humidity (84–91%), moderate evening humidity (40–75%), and weekly rainfall of 24–31 mm.

Logistic and Gompertz non-linear models effectively described disease progression, explaining 97–99% of the temporal variation, and enabled estimation of biological parameters such as infection rate and maximum disease severity. Peak disease severity occurred during the fourth–fifth week after backward pruning and the fifth–sixth week after fore pruning, indicating critical windows for disease management, calls for implementing prophylactic strategies. Area under Disease Progressive Curve (AUDPC) for Logistic and Gompertz for both backward and fore pruning data were also constructed to reorient the management strategy based on the population dynamics.

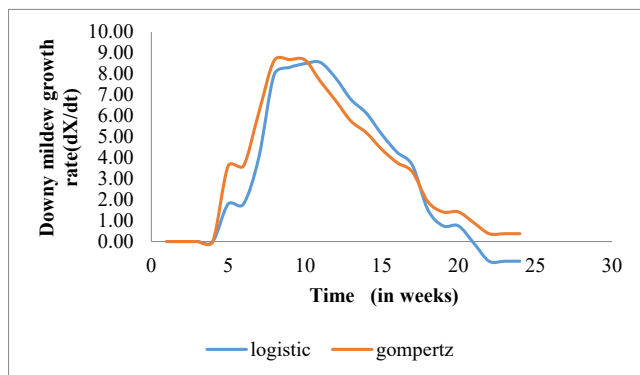


Fig. 8. Graphical Representation of Non-linear models

Details can be perused at (Venugopalan *et al.* 2015)

Non-Linear growth models for incidence of Powdery Mildew in early flowering variety of Mango (Lazzat Baksh)

This study investigated the incidence and temporal development of powdery mildew caused by *Oidium mangiferae* in the early-flowering mango cultivar Lazzat Baksh (*Mangifera indica* L.) over two growing seasons (2021–23). Secondary disease observation data recorded at ICAR–IIHR, Hesaraghatta, Bengaluru were

used to develop non-linear disease progression models using Logistic and Gompertz functions. Both models adequately described disease dynamics; however, the Logistic model showed superior performance with higher adjusted R^2 (0.95) and lower mean squared error (8.32). Pooling data across seasons enhanced model stability and predictive accuracy. Peak disease severity occurred during the first to second week, indicating a critical window for effective disease management and timely control measures.

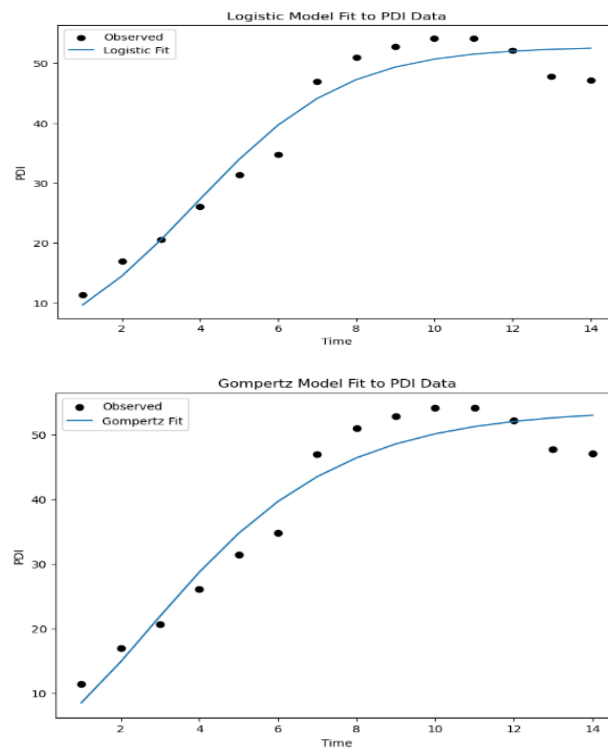


Fig. 9. Graphical Representation of Non-linear models

Details can be perused at (Pallavi *et al.* 2025)

7. FUTURE PROSPECTS

Future research studies should extend beyond deterministic non-linear models by incorporating stochastic differential equation-based approaches to account for environmental variability and temporal correlation in longitudinal disease data. In parallel, machine learning and deep learning techniques such as ensemble learning and recurrent neural networks can be explored to capture complex, non-linear interactions among weather variables and disease dynamics, thereby improving predictive accuracy and enabling real-time disease forecasting and decision-support systems.

8. NATURAL RESOURCE MANAGEMENT

8.1 Non-destructive Carbon Sequestration Estimation in Grafted Mango Using Non-linear Allometric Models

This study developed species-specific allometric equations based on non-linear regression methods for estimating biomass and carbon sequestration in grafted mango trees, overcoming the limitations of forest-based DBH models for orchard systems. Destructive sampling across age groups showed that stems and primary branches contributed nearly 50% of above-ground biomass, while roots accounted for about 22%. Biomass expansion factors varied with age and stabilized after 20 years. A power model using primary branch girth and number of primary branches best predicted above-ground biomass ($R^2 = 0.971$). Application of the model estimated 9.913 Mt of carbon sequestration across the Konkan mango belt (106,210 ha).

Further details can be seen at

(Ganeshamurthy *et al.* 2016; Ganeshamurthy *et al.* 2019)

Future studies may integrate machine learning, deep learning, and large language models with remote sensing and field data to enhance biomass estimation, scalability, and decision-support in orchard carbon accounting. Expansion of artificial neural network (ANN) frameworks by incorporating additional leaf traits, genotypic variability, and environmental covariates could further improve prediction accuracy. Validation of ANN models across diverse locations, seasons, and growth stages will strengthen model robustness and generalizability. Integration of image-based inputs and automated data acquisition techniques can simplify non-destructive leaf area estimation. Advanced deep learning architectures may be explored to capture complex non-linear relationships more effectively, positioning ANN-based approaches as viable alternatives to conventional regression methods in horticultural research.

9. POST-HARVEST TECHNOLOGY

9.1 Response surface models in Aonla and Dragon fruit to increase input-use efficiency in relation to crop productivity

9.1.1 Optimization of factors influencing osmotic dehydration in Aonla (*Phyllanthus emblica* L.) segments in salt solution using response surface methodology

Aonla (*Phyllanthus emblica* L.) is an important medicinal fruit rich in ascorbic acid and bioactive compounds, widely valued for its nutritional and therapeutic properties. However, its highly perishable nature and astringent taste necessitate processing and value addition to extend shelf life and market availability. Dehydration is an effective preservation method, with osmotic dehydration serving as a preferred pre-treatment due to its ability to retain nutritional and sensory quality. The efficiency of osmotic dehydration depends on process variables such as solute concentration, temperature, and time. Optimizing these parameters is essential to enhance water loss while minimizing solid gain without compromising product quality. Response Surface Methodology provides a systematic approach for achieving optimal processing conditions.

Optimization of process parameters is essential in food processing to develop highly acceptable products. In this study, mass transfer kinetics during osmotic dehydration in Aonla segments in salt solution were evaluated using a Box–Behnken Design (BBD) under the response surface methodology (RSM) framework. Salt concentration (2–6%), process temperature (45–55 °C), and process time (60–180 min) were selected as independent variables, while water loss (WL%) and solids gain (SG%) served as response parameters. Model validation was performed through experiments at optimum conditions. The optimized process conditions were 5.02% salt concentration, 54.8 °C temperature, and 60.64 min, resulting in maximum water loss (6.42%) and minimum solids gain (1.09%).

Optimization of factors influencing osmotic dehydration of aonla segments

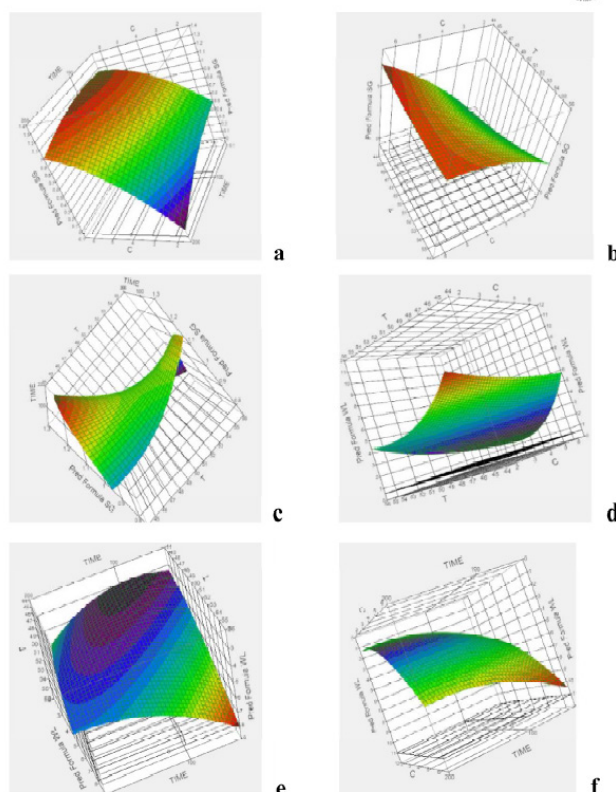


Fig. 4 : Illustration of profile for the quadratic response surface plot in the optimization of two variables for WL and SG

Fig. 10. Optimization of factors influencing osmotic dehydration of Aonla Segments

9.1.2 Optimization of fertilizer doses for increased fruit yield of Dragon fruit using response surface methodology

Response Surface Methodology (RSM) was used to predict dragon fruit yield responses under varying nutrient regimes using a Box–Behnken design. Second-order polynomial models were developed for the response variable (fruit yield), and the derived equations were used to identify optimal N:P:K fertilizer combinations for maximizing yield. The results indicate that the optimum doses of N:P₂O₅:K₂O for attaining higher yields of white (25.5 t ha⁻¹) and purple-red (35.6 t ha⁻¹) dragon fruit cultivars were 400:300:650 and 700:400:350 g per pillar per year, respectively. The non-significant lack-of-fit F-test ($P > 0.05$) demonstrates the adequacy and reliability of the developed models. Furthermore, graphical analyses combined with RSM proved effective in identifying optimal nutrient levels that are experimentally verifiable and reproducible.

Overall, the findings offer valuable guidance for efficient nutrient management in dragon fruit cultivation, which can enhance productivity, promote expansion of the cultivation area, and ultimately improve farm income

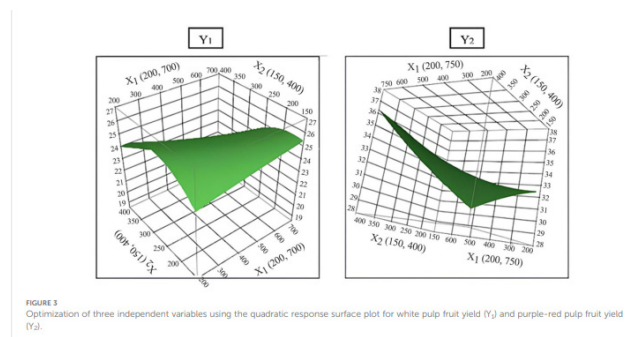


Fig. 11. Optimization of three independent variables using the quadratic response surface plot for white pulp fruit yield (Y₁) and purple-red pulp fruit yield (Y₂)

9.1.3 Optimization of process variables viz., temperature and time for vacuum concentration of dragon fruit juice, it's effect on TSS, water activity, total phenol content and antioxidant activity

This study represents the pioneer investigation into the vacuum concentration process applied to dragon fruit juice, as per the available literature. It investigates the influence of temperature and processing time on the vacuum concentration of dragon fruit juice in rotary evaporator. Through systematic experimentation and statistical analysis by Response surface methodology (RSM), using Central composite design (CCD). Key parameters such as total soluble solids (TSS), water activity, total betalain content, and antioxidant activity were evaluated. Results indicate significant effects of temperature and time on these parameters, with optimal conditions identified for maximizing TSS, betalain content, and antioxidant activity while minimizing water activity. Statistical modelling confirms the reliability of the findings, offering practical insights for optimizing the vacuum concentration process of dragon fruit juice to enhance its quality and handling efficiency.

Further details can be seen at

Karunakaran *et al.* 2025

Rosemary *et al.* 2025

Sujayashree *et al.* 2022

10. FUTURE STUDIES IN MASS TRANSFER KINETICS MODELLING

10.1 Role of ANN Models in Osmotic Dehydration and Mass Transfer Kinetics

Process variables in any osmotic dehydration studies of post-harvest products are naturally non-linearly related among themselves. Artificial neural network (ANN) models are increasingly used to describe osmotic dehydration and mass transfer kinetics in horticultural produce due to their ability to model complex nonlinear interactions among process variables. ANNs effectively predict moisture loss, solid gain, water activity, and quality attributes using inputs such as solution concentration, temperature, immersion time, and sample geometry. Unlike conventional kinetic models, ANN approaches do not require prior assumptions about mass transfer mechanisms. Their superior predictive accuracy supports process optimization, quality retention, and scale-up. Integration of ANN models with optimization techniques further enhances dehydration efficiency and process control in post-harvest systems.

11. CONCLUSION

Statistical techniques are central to horticultural research, enabling robust analysis of complex research issues entailing multi-disciplinary fields such as crop improvement, production, protection, post-harvest management, climate resilience, and socio-economic studies. In crop improvement programs, both conventional and molecular breeding approaches rely heavily on statistical tools for systematic germplasm evaluation, estimation of genetic parameters, assessment of genetic diversity, stability analysis, and construction of selection indices for identifying superior genotypes across environments. In molecular breeding, statistical and bioinformatics methods support quantitative trait locus (QTL) mapping, genome-wide association studies (GWAS), marker–trait association analysis, genomic selection, and integration of multi-omics datasets to accelerate genetic gain. Sustainable horticultural technologies increasingly rely on linear and nonlinear models, response surface methodology, and predictive analytics to optimize soil, seed, nutrient, and water use efficiency, including applications in perennial systems and off-season cultivation.

Statistical modelling also underpins integrated pest and disease management, post-harvest process optimization, and climate-resilient horticulture through ecological models, nonparametric methods, growth curve analysis, and stochastic approaches. In extension and agricultural economics, statistical and econometric techniques support technology evaluation, market intelligence, price forecasting, and research prioritization. As an interface, suitable R codes were developed (Venugopalan and Sisira, 2025).

Future research should focus on capturing non-linear functional relation among the biological traits through hybrid statistical–AI frameworks integrating machine learning, deep learning, and artificial neural networks (ANNs) with advanced architectures such as convolutional neural networks (CNNs) for image-based phenotyping, disease and quality assessment; graph neural networks (GNNs) for modelling genotype–phenotype–environment interactions, gene networks, and agro-ecosystem relationships; and large language models (LLMs) for decision-support systems, knowledge synthesis, advisory services, and automated interpretation of experimental and extension data. Coupled with high-throughput phenotyping, remote sensing, and simulation models, these approaches will enhance prediction accuracy, scalability, and real-time decision-making in precision and sustainable horticulture.

ACKNOWLEDGEMENTS

Greatly acknowledge with remembrance, the motivation and guidance of Prof. (Late). Prem Narian to pursue scientific research with a commitment to contribute to the statistical science as applied in various areas of agricultural science, during my Master's and Doctoral degree period (1991–1996) at IASRI, New Delhi.

REFERENCES

- Ganeshamurthy A N, Ravindra V and Rupa T R. (2019). Carbon Sequestration Potential of Mango Orchards in India. *Current Science* 117(12): 2006–13. <https://www.jstor.org/stable/27138603>.
- Ganeshamurthy A N, Ravindra V, Venugopalan R, Mathiazhagan M and Bhat R M. (2016). Biomass Distribution and Development of Allometric Equations for Non-Destructive Estimation of Carbon Sequestration in Grafted Mango Trees, *Journal of Agricultural Science* 8(8): ISSN 1916-9752. <http://dx.doi.org/10.5539/jas.v8n8p201>

- Hanumanthaiah R, Venugopalan R, Padmini K and Yogeesh K J. (2018). Yield Prediction in Brinjal (*Solanum melongena* CV MAHYCO-11) Across Different Growth Stages Using ANN Models. *International journal of Current Microbiology Applied Science* 7(4):803-810.
- Jalikop S H, Venugopalan R and Kumar, R. (2010). Association of fruit traits and aril browning in pomegranate (*Punica granatum* L.), *Euphytica*, **174**, 137-141, DOI: 10.1007/s10681-010-0158-3
- Karunakaran G, Ramachandran S, Venugopalan R, Selladurai R, Arivalagan M, Srivastava AK, Shilpa K G and Ruchitha T. (2025). Optimization of fertilizer doses for increased fruit yield of dragon fruit adopting response surface methodology and Box-Behnken design. *Frontier Plant Science* 16:1661974. DOI: 10.3389/fpls.2025.1661974
- Nassar R and Huehn M. (1987). Studies on estimation of phenotypic stability; Tests of significance for non-parametric measures of phenotypic stability. *Biometrics* **43**, 45-53.
- Sujayasree O J, Tiwari R B, Venugopalan R, Narayana C K, Bhuvaneshwari S, Ranjitha K, Oberoi H S, Shamina A, Sakthivel T and Nayaka V S K. (2022). Optimization of factors influencing osmotic dehydration of aonla (*Phyllanthus emblica* L.) segments in salt solution using response surface methodology, *Journal of Horticultural Sciences* **17**(2), 397-410.
- Nair S A, Sujatha K and Venugopalan, R. (2011). Influence of pruning time on enhancing the yield and quality of *Jasminum sambac* flowers during off-season. *The Indian Journal of Agricultural Sciences*, **79**(11). <https://epubs.icar.org.in/index.php/IJAgS/article/view/2578>
- Pallavi, R Venugopalan, S Sriram, Ranganath, H K, P Sisira. (2025). Non-Linear growth models for incidence of Powdery Mildew in early flowering variety of Mango (Lazzat Baksh), *The Indian Journal of Agricultural Sciences*. [communicated].
- Paul R K and Bhar L M. (2011). M-estimation in block design. *Journal of Indian Society of Agricultural Statistics*, **65**(3), 323-330.
- R Sukanya, R Venugopalan, K. Padmini, Ranganath. H. K, and Sisira P. (2025). Application of MANOVA statistical approach in Pumpkin (*Cucurbita moschata*) for identifying best treatment simultaneously based on several traits, *Journal of Horticultural Sciences* [communicated]
- Rosemary M X, Suresh G J, Venugopalan R. (2025). Optimization of process variables viz., temperature and time for vacuum concentration of pitahaya (*Hylocereus polyrhizus*) fruit juice, its effect on total soluble solids, water activity, total betalain content and antioxidant activity. *Discover Foods* **5**, 53. <https://doi.org/10.1007/s44187-025-00313-w>
- Thennarasu K. (1995). On certain non-parametric procedures for studying genotype-environment interactions and yield stability. *Indian Journal of Genetics* **60**, 433-439.
- Venugopalan R, Kumar R and Sisira P. (2025). On Non- Parametric approach for multi-traits based crop stability assessment in China Aster (*Callistephus chinensis* L.). *Journal of Horticultural Sciences*, **20**(2). <https://doi.org/10.24154/jhs.v20i2.4146> .
- Venugopalan R, Kurian R M, Chaithra M and Sisira P. (2023). An alternate statistical method for dealing outliers in perennial crop experiment. *Journal of Horticultural Sciences*, **18**(2).479-479.
- Venugopalan R, Pitchaimuthu M and Chaithra M. (2020). Non-Parametric Stability Approach for Horticultural Crop Varietal Release. *Journal of Indian Society of Agricultural Statistics*, **74**(1):73-80.
- Venugopalan R, Swamy K R M and Chandraprakash M K. (2008). Augmented BIB design: An alternative statistical design in germplasm evaluation trials. *Journal of Horticultural Sciences*, **3**(1), 84-88.
- Venugopalan R and Manjunath B L. (2019). Application of Robust ANOVA methods in papaya having outlier data. *Journal of Indian Society of Agricultural Statistics*, **73**(2), 137-140.
- Venugopalan R. (2010). Application of statistical principles for evolving crop logging models in banana (Cv. Ney Poovan). *Tropical Agriculture*, **87**(1), pp. 29-32
- Venugopalan R and N. Vijay. (2015). Nonlinear Logistic Model for Describing Downy Mildew Incidence in Grapes. *Journal of Indian Society* **69**(1):19-25.
- Venugopalan R and Sisira P (2025). Teaching Manual for “Application of basic and advanced statistical methods for horticultural crops research: R-interface”, Practical Manual_IHR-TM-8/2025., 149p.
- Venugopalan R and Sisira P (2025). Statistical Modelling and Machine Learning and Deep Learning Methods”, Practical Manual_IHR-TM-9/2025., 22p.
- Venugopalan R and Sisira P (2025). Biometrical Genetics: Basics and Advanced., ICAR-IIHR, Practical Manual_IHR-TM-10/2025., 49p.
- Venugopalan R and Sisira P (2025). Practical manual for Gene action studies using R for horticultural crop research., TM-5/2025., 30p.